

# Ex Ante and Ex Post Evaluations: Two Sides of the Same Coin?

## The Case of Gender Mainstreaming in EU Research Policy

Lut Mergaert and Rachel Minto\*

*This article engages with two themes of contemporary EU governance: the role of evaluations within an effective and coherent policy-making process and the EU's constitutionalised commitment to promoting gender equality in all its activities (Article 8 TFEU). It focuses on the interface between ex ante and ex post evaluation and the contribution of evaluations to policy learning, with particular attention to the promotion of gender equality. A case study approach is followed, with EU research policy as the object of analysis.*

### I. Introduction

At the time of writing, the European Commission is developing a renewed set of Guidelines for Evaluation, as part of its Smart Regulation agenda.<sup>1</sup> Key to an efficient, effective and coherent policy cycle is that the *ex ante* and *ex post* components of evaluation are linked.<sup>2</sup> However, this is no easy task. Even scholarship on evaluation is largely separated into two camps, with few EU scholars addressing both *ex post* and *ex ante*. This article contributes to bridging this gap. The overarching question is whether *ex ante* and *ex post* evaluations are, indeed, two sides of the same coin.

To explore this link, this article engages another horizontal theme within European governance,

namely the EU's constitutionalised commitment to promoting gender equality in all its activities (Article 8 TFEU), known as "gender mainstreaming". Given its cross-sectoral applicability, the procedural requirements attached to gender mainstreaming are thus relevant in all EU policy, including regulatory policy. Whilst gender mainstreaming implementation has enjoyed mixed success at best,<sup>3</sup> this article focuses on EU research policy as an area which provides a more positive example of gender mainstreaming in the EU<sup>4</sup> and which, given the expenditure attached to it, has a long-established culture of evaluation.<sup>5</sup> Policy-making in this area works on a cyclical basis, around the reiterations of the Research and Technological Development (RTD) Framework Programmes (FPs) (the key instrument in this area), for

\* Lut Mergaert is a Research Director at Yellow Window and holds a PhD in Management Sciences from Radboud University Nijmegen; Rachel Minto is a Research Associate at Cardiff School of Law and Politics and holds a PhD in Politics from the University of Bristol. Rachel Minto would like to thank the European Research Council for funding her research for this article, under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 313642-LASI ("Law, science and interests in European policy-making").

1 See European Commission, "Consultation on the draft Commission Evaluation Policy Guidelines", 10 April 2014, available on the Internet at <[http://ec.europa.eu/smart-regulation/evaluation/consultation/index\\_en.htm](http://ec.europa.eu/smart-regulation/evaluation/consultation/index_en.htm)> (last accessed on 23 May 2014). The consultation period ended on 25 February 2014.

2 Vítor Oliveira and Paulo Pinho, "Bridging the gap between planning evaluation and programme evaluation: The contribution of the PPR methodology", 17(3) *Evaluation* (2011), pp. 293 *et seq.*

3 Emilie Hafner-Burton and Mark Pollack, "Mainstreaming Gender in the European Union: Getting the Incentives Right", 7(1) *Comparative European Politics* 2009, pp. 114 *et seq.*; Emanuela Lombardo and Petra Meier, "Gender Mainstreaming in the EU: Incorporating a Feminist Reading?" 13(2) *European Journal of Women's Studies* (2006), pp. 151 *et seq.*; Teresa Rees, "Reflections on the uneven development of gender mainstreaming in Europe", 7 *International Feminist Journal of Politics* (2005), pp. 555 *et seq.*; Maria Stratigaki, "Gender Mainstreaming vs Positive Action: An Ongoing Conflict in EU Gender Equality Policy", 12(2) *European Journal of Women's Studies* (2005), pp. 165 *et seq.*

4 Emilia Hafner-Burton and Mark Pollack (2009), *ibid*; Mark Pollack and Emilie Hafner-Burton, "Mainstreaming gender in the European Union", 7(3) *Journal of European Public Policy* (2000), pp. 432 *et seq.*; Teresa Rees, "Mainstreaming Gender Equality in Science in the European Union: the 'ETAN Report'", 13 *Gender and Education* 2001, pp. 243 *et seq.*

5 Steven Højlund, "Evaluation in the European Commission - For accountability or learning?", in this Symposium. For an overview of the history of evaluation in DG Research, please see Erik Arnold, Bea Mahieu, James Stroyan, David Campbell, Malin Carlberg, Flora Giaracca, Andrej Horvath, Zsuzsa Jávorka, Paula Kneee, Ingeborg Meijer, Sabeen Sidiqi and Caroline Wagner, "Understanding the Long Term Impact of the Framework Programme", Final Report To the European Commission DG Research by the European Policy Evaluation Consortium (EPEC) 2011, at pp. 5-11.

which there are multiple, pre-determined evaluation points. Throughout this iterative policy process, different types of evaluations have been undertaken: *ex ante* and *ex post*; general and gender-specific. Investigating gender equality policy learning in EU research policy therefore provides a particularly suitable case to explore the contribution of evaluation to policy learning in EU decision-making.

Analysis of the relationship between *ex ante* and *ex post* evaluations demands attention to two specific research questions. Firstly, how and to what extent are *ex ante* and *ex post* evaluations in EU research policy linked coherently and effectively with respect to gender equality? Secondly, what role has (*ex post* and *ex ante*/general and gender-specific) evaluation played in the process of gender mainstreaming in EU research policy? Research was pursued through a text-based analysis of selected evaluation documents, identified as potential key resources for use during the transitional moments between policy cycles.

Four sections and a conclusion follow this introduction. Section II provides the theoretical underpinning for analysis, locating this empirical study with respect to scholarship on evidence-based policy-making and the linking of *ex ante* and *ex post* evaluations. Section III presents EU research policy and the history of gender mainstreaming throughout the multiple iterations of the FPs. Section IV introduces the evolving system of evaluation in EU research policy, presenting the evaluation documents selected for analysis. The findings in relation to the research questions are shared in Section V, organised as follows: 1) mis/matching evaluations; 2) non-linear policy learning; 3) differences between different evaluation types; and 4) the importance of institutional factors. Section VI concludes the article by returning to the principal focus of investigation, namely whether *ex ante* and *ex post* evaluations are indeed linked in terms of promoting gender equality in EU research policy. The article argues that there is little evidence to suggest the coherence and effectiveness of this *ex ante/ex post* link. In terms of policy learning, whilst there are examples of (selective) coherence between evaluations, broadly speaking the formative potential of evaluations is hampered by procedural, political and institutional factors. The article offers some suggestions to strengthen the link between *ex ante* and *ex post* evaluations in EU policy-making, with respect to gender equality in particular, as well as more broadly.

## II. Linking *ex ante* and *ex post* Evaluations as Part of a Coherent Policy Cycle

The link between *ex ante* and *ex post* evaluations in the policy-making process remains remarkably under-theorised.<sup>6</sup> Indeed, the scholarship on policy evaluation in the EU is predominantly separated into two camps which attend to *ex ante* evaluation (namely the Commission's Integrated Impact Assessment system) and *ex post* evaluation (with a particular methodological focus) largely in isolation from each other. As such, in terms of a theoretical underpinning for this empirical study, it was necessary to hunt more broadly beyond "evaluation scholarship" *per se*. From the selection of suitable approaches outlined in the Introduction to this Special Issue, the literature on evidence-based policy-making (EBPM) and policy learning provide some points of reference around which to anchor analysis.

In particular, the policy cycle approach underpinning EBPM is useful for the purposes of this research. This more rationalist account of policy-making distinguishes between various stages of the policy process, commonly: agenda setting, formulation, decision-making, implementation, and evaluation.<sup>7</sup> In accordance with EBPM, the use of knowledge gained through *ex post* evaluations in *ex ante* evaluations is a crucial link in the policy cycle, as such evidence informs policy-making and contributes to policy learning as part of effective governance.<sup>8</sup> Such an understanding of the policy process has, however, been criticised,<sup>9</sup> in part because the theoretical account of the policy cycle – with distinct, neatly compartmentalised stages – is not reflected in practice. Whilst not unsympathetic to such critiques, for the purposes of the current analysis it was considered to provide a useful framework nonetheless. Notably, the EBPM approach to policy-making was adopted in the UK in the late-1990s, with other Organisation for Econom-

6 Stijn Smismans, "Policy evaluation in the EU: the challenges of linking *ex ante* and *ex post* appraisal", in this Symposium.

7 See, e.g. Esther Versluis, Mendeltje van Keulen and Paul Stephenson, *Analyzing the European Union Policy Process* (Basingstoke, New York: Palgrave Macmillan, 2011).

8 Ian Sanderson, "Evaluation, policy learning and evidence-based policy making", 80(1) *Public Administration* 2002, pp. 1 *et seq.*

9 For a discussion of the advantages and disadvantages of this "stages approach", see. Esther Versluis, Mendeltje van Keulen and Paul Stephenson (2011), *supra*, note 7, at pp. 21-24.

ic Co-operation and Development (OECD) countries and the EU following suit.<sup>10</sup> Indeed, with particular reference to the linking of *ex post* and *ex ante* evaluations, this linear understanding of the policy process can be seen in EU policy texts, e.g. the annual reports from the Impact Assessment Board (IAB) have made repeated references to the use of results from *ex post* evaluations in the (*ex ante*) IA process.<sup>11</sup>

Regarding gender mainstreaming in the practice of evaluation, in terms of effective governance for the promotion of gender equality, one would expect to see a number of clear gender equality objectives traced over the multiple iterations of the Framework Programmes. As part of a process of policy learning, the knowledge gained from *ex post* evaluations would be used to inform discussions around policy design in the *ex ante* evaluations. It is through such an objective-orientated and evidence-based approach that, in theory, policy-making can advance the EU's constitutionalised commitment to gender equality (Article 8 TFEU).

Evaluation scholars have, however, highlighted potential challenges attached to the linking of *ex*

*ante/ex post* evaluations. These challenges relate to the largely independent development of these evaluation practices, which may (partly) explain the difficulty of one type feeding into another.<sup>12</sup> They have “different evaluation scopes and contexts, different relationships between research and professional practice, and different timings”.<sup>13</sup> Notably, ensuring coherence across evaluation exercises and maximising their formative potential is intimately connected to the usability and actual use of evaluation outputs.<sup>14</sup> In particular the timing and nature of the release of evaluation outputs are pertinent in terms of maximising evaluation use, and should (thus) be embedded in decision-making cycles.<sup>15</sup>

These challenges may be heightened in the area of EU research policy, given the complexity and scope of the EU's RTD Framework Programmes, and the multiplicity of pursued objectives and related stakeholders. This complexity has implications in terms of evaluation: “In the evaluation of complex issues, judgment comes from many independent sources, through many technical means, using multiple criteria embedded in different value systems”.<sup>16</sup> Furthermore, this complexity leaves the consideration of gender equality in a particularly vulnerable position as unless it is explicitly included in the scope of an evaluation there is no guarantee that it will be considered.<sup>17</sup> Indeed, “[t]he all-purpose evaluation is a myth ... Not even a well-planned study will provide information on all the questions for all the people involved”.<sup>18</sup> This indicates that, as part of an effective evaluation process that seeks to contribute to gender equality, one would expect gender equality objectives to be highlighted as a distinct concern to be addressed.

Therefore, based on existing scholarship, it can be assumed that there are clear challenges to ensuring coherent policy learning over the multiple iterations of the EU's Framework Programmes. This article focuses squarely upon this subject, providing new empirical research on the practice of evaluation in the EU, as part of the EU's Smart Regulation agenda.

### III. Setting the scene: EU Research Policy and Gender Mainstreaming

From more humble beginnings in the early days of European cooperation, EU research policy developed into a well-established area of EU activity, mobilising a not insignificant portion of the EU budget. With

10 Linda Courtenay Botterill and Andrew Hindmoor, “Turtles all the way down: bounded rationality in an evidence-based age”, 33(5) *Policy Studies* (2012), pp. 367, et seq., at p. 369 which includes reference to Kai Böhme, “Much Ado about Evidence: Reflections from Policy Making in the European Union”, 3(1) *Planning Theory and Practice* (2002), pp. 98 et seq.

11 See the Impact Assessment Board Reports for 2009, 2010, 2011, 2012 and 2013.

12 Vítor Oliveira and Paulo Pinho (2011), supra, note 2.

13 Vítor Oliveira and Paulo Pinho (2011), supra, note 2, at p. 305.

14 Vítor Oliveira and Paulo Pinho (2011), supra, note 2; Raynald Pineault, Paul Lamarche, Marie-Dominique Beaulieu, et al., “Conceptual and Methodological Challenges in Producing Research Syntheses for Decision-and Policy-Making: An Illustrative Case in Primary Healthcare”, 16(2) *Evaluation* (2010), pp. 137 et seq.; Murray Saunders, “The use and usability of evaluation outputs: A social practice approach”, 18(4) *Evaluation* (2012), pp. 421 et seq.

15 Saunders (2012), *ibid*, at p. 428.

16 Bojan Radej, “Synthesis in policy impact assessment”, 17(2) *Evaluation* (2011), pp. 133 et seq., at p. 134.

17 Juan Andres Ligerio Lasa, Julia Espinosa Fajardo, Carmen Mormeneo Cortes, María Bustelo Ruesta, *Making Evaluation Sensitive to Gender and Human Rights. Different Approaches*, (Spanish Ministry of Foreign Affairs and Cooperation, Secretary of State for International Cooperation and for Ibero-America, General Secretary of International Cooperation for Development Affairs and Cooperation Secretary of State for International Cooperation and for Ibero-America General Secretary of International Cooperation for Development, 2014), at p. 17.

18 Carol H. Weiss, *Evaluation. Methods for Studying Programs and Policies*, Second Edition (Upper Saddle River (NJ): Prentice Hall, 1998), at p. 33, quoted in Juan Andres Ligerio Lasa, Julia Espinosa Fajardo, Carmen Mormeneo Cortes, María Bustelo Ruesta (2014), supra, note 17, at p. 73.

its home in DG Research and Innovation (DG RTD), central to EU research activity are the RTD Framework Programmes. These multiannual funding programmes set the objectives and priorities for European research activity which, since the 1997 Treaty of Amsterdam, have been agreed according to the ordinary legislative procedure. From the first five-year programme (FP1) to the current seven-year programme (Horizon 2020), the cyclical policy-making process has been repeated seven times. However, it was not until FP5 (1998-2002) that gender was included, so it is here that the analysis begins: FP5 (1998-2002) as the starting point, into FP6 (2002-2006), FP7 (2007-2013), and up to Horizon 2020 (2014-2020); thus encompassing three revisions of the regulatory framework. These three “transition points” and the *ex post* and *ex ante* evaluations captured by them, are the focus of this work.

Since its adoption in EU research policy, the implementation of gender mainstreaming has been marked by ups and downs.<sup>19</sup> Gender mainstreaming was adopted by the Commission in 1996,<sup>20</sup> defined as “not restricting efforts to promote equality to the implementation of specific measures to help women, but mobilising all general policies and measures specifically for the purpose of achieving equality” (emphasis in original). The 1997 Treaty of Amsterdam then formalised this commitment, enshrining in the Treaties the elimination of inequality and the promotion of equality between women and men as an aim, horizontally applicable across Community (now Union) activities (Article 3(2) EC [now Article 8 TFEU]); making it a duty for civil servants in the Commission to integrate the gender perspective in all EU policies.

Adopted when launching FP5,<sup>21</sup> gender mainstreaming in EU research policy took off with conferences and expert groups, analyses to study the problems at stake, the definition of a dual objective (to enhance women’s participation and to address the gender dimension in research content), and an agenda for future work.<sup>22</sup> The next Framework Programme (FP6) was marked by a “strong thrust” forward, with legally fixed provisions.<sup>23</sup> Among other measures, FP6 introduced a specific instrument (the Gender Action Plan, or “GAP”) for integrating a gender perspective in the (biggest) funded projects, and gender monitoring studies were commissioned to follow up progress and to report on obstacles and challenges. In the course of FP6 and towards FP7, how-

ever, the support for gender mainstreaming in research policy waned.<sup>24</sup> This translated into a regression in terms of provisions in FP7: the GAPs were abandoned and no monitoring efforts were foreseen. Instead, a new initiative was developed (a toolkit on gender in EU-funded research) and training was offered to the research community on how to integrate a gender perspective in research. In response, stakeholders united and exerted pressure on the Commission, such that it seems that gender mainstreaming and gender equality are more firmly on the research policy agenda again under Horizon 2020.

#### IV. *Ex ante* and *ex post* Evaluations of EU Research Policy

EU research policy has a tradition of evaluation. Indeed, given the expenditure attached to research policy, DG RTD was one of the first to develop its evaluation function.<sup>25</sup> Over the years, EU research policy has developed considerably in both scale and scope (influencing the types of evaluation required) and the accompanying system of evaluation and monitoring has evolved in line with broader changes to the culture of evaluation in the Commission.<sup>26</sup> This

19 Lut Mergaert and Katlijn Demuyne, *The ups and downs of gender mainstreaming in the EU research policy - the gender toolkit and training activities in FP7* (Antwerp, Belgium: Policy Research Centre on Equal Opportunities, 2011).

20 Commission Communication on Incorporating equal opportunities for women and men into all Community policies and activities, COM (96) 67 final, at p. 2.

21 The Commission’s Communication on gender mainstreaming was published in 1996, when FP4 was running (1994-1998).

22 European Commission, *Stocktaking 10 years of “Women in Science” policy by the European Commission (1999-2009)*, (Luxembourg: Publications Office of the European Union, 2010), NB This is also text#9; Mergaert and Demuyne (2011), *supra*, note 19.

23 European Commission (2010), *ibid*, at p. 216. NB This is also text#9.

24 Lut Mergaert, “The Reality of Gender Mainstreaming Implementation. The Case of the EU Research Policy” (Doctoral dissertation, Radboud Universiteit Nijmegen, Nijmegen, 2012).

25 Steven Højlund (2014), *supra*, note 5.

26 Hilka Summa and Jacques Toulemonde, “Evaluation in the European Union: Addressing Complexity and Ambiguity”, in Jan-Eric Furubo, Ray. C. Rist and Rolf Sandahl (eds.), *International Atlas of Evaluation* (New Brunswick and London: Transaction Publishers, 2002), pp. 407 *et seq.*

For an overview of the history of evaluation in this policy area, see Eric Arnold, Bea Mahieu, James Stroyan, *et al.* *Understanding the Long Term Impact of the Framework Programme, Final Report To the European Commission DG Research* (European Policy Evaluation Consortium (EPEC), 2011), at pp. 25-27.

evolution presents particular challenges and opportunities as the period of analysis (1998-2014) captures definitive moments of transition. Following a review, ten texts were identified as key evaluations attached to the development of EU research policy (for a list of the selected texts, please see the annex). This corpus for analysis comprised both *ex ante* and *ex post* evaluations, including both general and gender-specific. The annual monitoring reports sit beyond the scope of the current project.

The *ex ante* texts identified for analysis were the two Integrated Impact Assessments (IIA) accompanying the transitions from FP6 to FP7, and from FP7 to Horizon 2020. The IIA was established in 2002,<sup>27</sup> becoming “fully operational” in 2004/5.<sup>28</sup> Prior to this, there was no formal system of *ex ante* assessment *per se*, although retrospective evaluations could inform subsequent FPs. As such, there was no IIA accompanying the development of FP6. The IIA replaced previously distinct impact assessment practices in the Commission (including for gender assessment),<sup>29</sup> as an instrument of the EU’s Smart Regulation agenda.<sup>30</sup>

Four *ex post*/retrospective evaluations were captured by the period of analysis. The first two were Five-Year Assessments (FYA). The one covering the period 1995-1999 (published in 2000) was available for the shift from FP5 to FP6 and the other, covering 1999-2003 (published in 2004), falls within the boundaries of the shift from FP6 to FP7, although they are not a perfect match. The remaining two retrospective evaluations were specifically attached to the FP (these took over from the FYA from FP6). They consisted of an external, *ex post* evaluation of FP6 (published in 2009) and an external, interim evaluation of FP7 (published in 2010).

In addition to these six general evaluations of the FPs, four gender-specific evaluation-related exercises were identified for analysis, all retrospective in nature. The first was the (retrospective) Gender Impact Assessment attached to FP5. It was released in 2001, near the end of the FP cycle. It was undertaken by external contractors and culminated in a synthesis report covering the respective specific programmes. Two gender-specific evaluation exercises were attached to FP6. The first was a gender equality report prepared by the Commission, released in October 2008. The second, released in May 2009, was the Gender Monitoring Studies report, comprising a synthesis report of key findings and recommendations of six studies carried out between 2004 and 2007 to monitor progress towards gender equality and gender relevance awareness. Finally, in 2010, the Commission released a stocktaking text regarding 10 years of “Women in Science” policy (1999-2009).

The ten selected texts do not produce a wholly orderly timeline of *ex post/ex ante* evaluations clustered around the moments of transition. Instead, due to their dates of release, they are sometimes out of sync with the review and reiteration of the next FP. However, the evaluation narrative that emerges does include both *ex ante* (two) and *ex post* (eight) evaluations, comprising general (six) and gender-specific (four) evaluations.

## V. The Reality of the Link between *ex ante* and *ex post* Evaluations

Tracing gender mainstreaming through the selected texts returns a number of notable findings. These are unpacked below around the following inter-related themes: 1) mis/matching evaluations; 2) policy learning; 3) evaluation types; and 4) institutional factors.

### 1. Mis/matching Evaluations

The first finding relates to the timing of the evaluations. As highlighted above, due to their dates of release, evaluation-related exercises are sometimes out of sync with the review and reiteration of the next FP. Indeed, the Gender Equality Report of FP6 (text#5), the general *ex post* evaluation of FP6 (text#6) and the Gender Equality Monitoring in FP6

27 Commission Communication on Impact Assessment, COM(2002) 276 final.

28 Norman Lee and Colin Kirkpatrick, “Evidence-based policy-making in Europe: an evaluation of European Commission integrated impact assessments”, 24(1) *Impact Assessment and Project Appraisal* (2006), pp. 23 *et seq.*, at p. 24.

29 Commission Communication on Impact Assessment, *supra*, note 27, at p. 3.

30 The principle of continuous assessments underpins the Smart Regulation agenda: “To ensure that EU action is effective, the Commission assesses the impact of policies, legislation, trade agreements and other measures at every stage - from planning to implementation and review.” See European Commission, “Smart Regulation”, 13 June 2013, available on the Internet at <[http://ec.europa.eu/smart-regulation/index\\_en.htm](http://ec.europa.eu/smart-regulation/index_en.htm)> (last accessed on 23 May 2014).

Report (text#7) were all published after the launch of FP7. As such, they could not contribute to the *ex ante* evaluation process for FP7 (text#8), undermining the formative potential of these retrospective exercises. In addition, the two FYAs both straddled two FPs: FYA (1995-1999) (text#1) covered parts of both FP4 and FP5, and FYA (1999-2003) (text#3) covered parts of both FP5 and FP6. These mismatches partly reflect the time-lag between policy decision, intervention and “results” in this policy area; as such, this “finding” may appear pedestrian. However, it has important implications for policy learning and the use of evaluation results, given that results from one cycle are not available until the next cycle is already underway, while they ought to be embedded in the decision-making cycle.<sup>31</sup>

This mismatch also posed methodological questions regarding where links between evaluations “ought” to be expected. It was decided to organise the texts chronologically, according to their date of release. However, there may also have been some internal awareness of the content of the evaluation-related exercises within the Commission prior to their official release.

As noted in Section II, the period of analysis captures definitive moments in the development of the monitoring and evaluation system. Of particular note is the introduction of the *ex ante* IIA exercise in the early 2000s, because it institutionalised a clear point in the policy cycle for an *ex ante* assessment that did not exist previously.

## 2. The Formative Effect of Evaluation Work: Policy Learning over the FP Cycles

As outlined above, the success of gender mainstreaming has varied in EU research policy from FP5 to Horizon 2020. Referring to one of the gender-specific evaluation-related exercises (text#9, p. 195), the Commission itself attributes this fact to political priorities: “One of the major problems encountered has been the discontinuity of political support. There have been moments in which a wave of interest has rapidly driven activities, interspersed with moments of stagnation when all good intentions were left behind”. This variation is reflected in how and the extent to which gender equality is addressed within the evaluation texts.

Looking across the general texts, broadly speaking, they address three objectives: 1) increasing the number of women in science; 2) ensuring the needs of both women and men are met through scientific research, i.e. through making research gender-sensitive; and 3) increasing the representation of women in decision-making around the EU’s research activities (for which a target of 40% was set in 1999).<sup>32</sup> However, there is no clear thread of policy learning from FP5 to Horizon 2020 that develops with respect to these objectives. Women’s participation in science receives the most consistent attention. As suggested by the Gender Monitoring Studies (text#7), this dimension is probably the easiest to grasp, and definitely easier to monitor than the extent to which research takes the gender dimension into account. The others are addressed to varying degrees.

Focusing specifically on the interface between *ex post* and *ex ante* evaluations, the transitions from FP6 to FP7, and then from FP7 to Horizon 2020 provide clear examples of loss of policy coherence, as gender equality objectives were lost in these *ex post/ex ante* gaps. Indeed, the FYA (1999-2003) released in December 2004 (text#3) presented a broad gender equality agenda, addressing the coordination of RTD with other socio-economic policies, including gender (p.iv); the inclusion of women in science (pp. 4 and 13) to make use of “a valuable human resource that the EU cannot afford to waste” and to avoid the “distortion of the relationship between science and society” (p. 4); and women’s participation in decision-making (p. 13). Notably, it stated that it was important to “develop awareness of the benefits of integrating gender into research activities” and that “[continued] vigilance is necessary to transform European culture to raise gender sensitivity in science, technology and innovation” (p. 13). It also asserted that the Commission was to approve “the science, technology and innovation indicators” that should involve *inter alia* indicators of female participation in RTD (pp. 17-18). This evaluation was released during FP6, at a time when gender mainstreaming was more firmly embedded within the FP. Importantly, the IIA for FP7 recognises the FYA (1999-2003) as a “key input” in the development of FP7 (p. A1 – 40). However, this

31 Murray Saunders (2012), *supra*, note 14, p. 428.

32 Commission of the European Communities, *Women and Science: mobilising women to enrich European research* (Brussels, 1999).

was not the case with respect to gender. In the main report of the IIA, the attention paid to gender was minimal; these elements of the FYA were not picked up save for reference to the (potential) contribution of a particular programme (“People”) to rendering scientific careers more attractive for European citizens, including women (p. 8). The agenda does pick up other elements in the annex; for example, there is reference to “ensuring that societal dimensions, including the gender dimension, are integrated in research output where appropriate” (p. 56 – Annex 1), as part of the consideration of “science in society” under the “Capacities” programme. This part of the annex was referenced in the corresponding part of the IIA; however, having not been brought into the main report itself, gender equality is clearly afforded a diminished status compared to the preceding FYA (1999-2003). Indeed, as outlined above, gender mainstreaming was not as prominent in FP7 compared to the early days of FP6. Notably, a key instrument (the GAP) was abandoned under FP7.

A similar loss in coherence is identifiable between the FP7 interim evaluation (text#8) and the IIA for Horizon 2020 (text#10). Indeed, in its final pages and taking a forward-looking approach under the heading of “unfinished business”, the FP7 interim evaluation explicitly argues that there is a “compelling case for much more pro-active approaches to a better gender balance in the FP” (p. 76); however, the subsequent IIA does not attend to this, instead focusing on gender in terms of administrative costs. In contrast, the gender dimension in Horizon 2020 seems more developed than its predecessor, FP7.

Despite these clear examples of incoherence between evaluations, there was some evidence that evaluation exercises “spoke” to each other. Beginning with a more positive example, there is evidence of coherence between the *ex post* evaluation of FP6 (text#6) and the interim evaluation of FP7 (text#8). As per the Decision establishing FP7 (No 1982/2006/EC), the *ex post* evaluation for FP6 (text#6) itself reiterated (p. 2) that it was intended to inform the interim evaluation of FP7 (text#8); and there are certain indications that this was the case with respect to gender equality. Indeed, in a brief subsection entitled “FP6 outcomes and recommendations”, which recalls the preceding retrospective eval-

uation with respect to gender, the interim evaluation of FP7 referred to the “glass ceiling” for female researchers, stating that it was another “important lesson” from FP6 (p. 46). It also noted that advice in the FP6 *ex post* evaluation to continue the GAPs in FP7 “was not followed” (p. 45) but that there had been follow-up in terms of collecting and monitoring sex disaggregated data in line with suggestions advanced in the *ex post* evaluation (p. 45).

There are also examples of selective coherence between evaluation exercises. One particularly striking example of this is where gender equality lost out to a competing evaluation finding. In this case, the *ex post* FP6 evaluation stated that the decision to abandon the GAPs under FP7 was an “unfortunate example of simplification” (p. 25). The need for simplification was, however, also a finding from evaluation (namely the FYA 1999-2003 (text#3, p.iii) and repeated in the *ex post* evaluation of FP6); and one which was, in this case, prioritised over promoting gender equality in the FPs.

Another example of selective coherence can be seen between the evaluation text and gender mainstreaming in the FP itself, with respect to the recommendations from the Gender Monitoring Studies under FP6 (culminating in text#7). The need for capacity building was highlighted, and there was also a recommendation to keep the GAPs. However, whilst the former was done, the latter was ignored.

Given the complexity of EU research policy and the resulting challenges for evaluators,<sup>33</sup> it is not wholly surprising if policy concerns are lost between evaluation exercises. Also, the observed “mismatch” in terms of the timely availability of evaluation outputs (outlined in part i) may well act as a procedural restriction to policy coherence and policy learning (discussed in part ii). However, the fluctuation observed with respect to gender equality suggests that there are other procedural challenges, a point which is explored next.

### 3. Different Evaluation Types

In terms of the different types of evaluation (*ex post* and *ex ante*/general and gender-specific), the analysis points towards two research findings. The first concerns the difference between *ex ante* evaluations (namely the IIA) and *ex post* evaluations. As outlined in part 2 above, there was a noticeable loss in coher-

33 Bojan Radej (2001), *supra*, note 16, at p. 134.

ence between the IIAs for FP7 (text#4, particularly the main report) and Horizon 2020 (text#10) and the corresponding retrospective evaluations that preceded them. So, whilst welcoming the introduction of a clear *ex ante* evaluation point with requirements as to what should be addressed in such exercises, in practice, the link between IIAs and *ex post* evaluations is problematic in terms of realising the formative potential of evaluations. Indeed, it was striking that there was relatively poor attention to gender concerns in the two IIAs. This recalls existing research that found social issues in general, and gender issues in particular, to be under-addressed in the Commission's IIA exercises.<sup>34</sup>

A second point concerns the difference between general and gender-specific evaluation-related exercises. It is not uncommon for specifically-focused evaluations to be undertaken in the context of a broader policy area. Such evaluations may mobilise specific expertise to support the development/refinement of particular elements of a wider intervention; they also ensure that these elements (which may be overlooked in general evaluations) are addressed.<sup>35</sup> However, as part of a coherent and effective system of evaluation, it would be reasonable to expect that knowledge gained from these specifically-focused exercises would feed into the general evaluation process, thus capitalising on their formative potential. With respect to gender-specific evaluation exercises in EU research policy, as has been acknowledged previously,<sup>36</sup> this was not the case. Notably, there was relatively little visibility afforded to the gender-specific evaluation-related exercises. For example, whilst a number of individual reports prepared as part of the FP6 Gender Monitoring Studies had been released in advance of the publication of the synthesis report (text#7) and were therefore available for the FP6 *ex post* evaluation report (text#6), the latter does not make reference to these available individual reports and neither are they listed among this study's sources. However, the Commission's self-assessment about women's participation in FP6 (text#5) was considered. Commission-authored texts are not systematically included in general evaluation exercises, though, as the Commission's 2010 stocktaking text regarding 10 years of "Women in Science" policy (1999-2009) (text#9) was not referred to in the IIA for Horizon 2020 (text#10), published a year later.

In addition to political factors which push gender considerations in and out of favour and procedural

factors that undermine policy learning (highlighted above), institutional factors also impact upon evaluation (discussed below).

#### 4. Institutional Factors

A full engagement with institutional factors is beyond this article's scope, although they are certainly relevant to the question of evaluation so require some attention. Indeed, references within the evaluation-related exercises themselves indicate the existence of particular challenges attached to specific institutional factors. The first of these relates to the relatively high turnover of staff within the Commission (text#7, p. 7). There are also others, and the Commission itself recognises the problems, admitting that: "The [gender monitoring] studies encountered difficulties in collecting data from the Commission services, due in particular to the lack of timely and adequate information systems. Problems were also experienced due to structural reorganisations in the Commission, as well as personnel changes" (text#9, p. 86). As such, when it comes to the development of an institutional consciousness and policy learning from one cycle to the next, these are undermined by the "liquid nature of the bureaucracy",<sup>37</sup> which sees the frequent displacement/replacement of staff across the Commission as well as the re-organisation of DGs and Units.

Another factor is the apparent institutional importance (or lack thereof) attached to both gender mainstreaming and evaluation, as systemic approaches that ultimately aim at improving the quality and relevance of EU policies. Indeed, it is worth noting that the coordination of the evaluation function in the Commission has been placed at the level of the Secretariat-General which oversees all evaluations (that is, evaluations "strictu sensu") done by different Directorates-General. By contrast, the gender mainstreaming coordination function is situated in a unit

34 Lut Mergaert and Nathalie Wuiame, *Report on Institutional Capacity for Gender Mainstreaming in the European Commission*. Report from a study for the European Institute for Gender Equality (unpublished work, 2013).

35 Carol H. Weiss (1998), *supra*, note 18, at p. 33, quoted in Juan Andres Ligerio Lasa, Julia Espinosa Fajardo, Carmen Mormeneo Cortes, María Bustelo Ruesta (2014), *supra*, note 17, at p. 73.

36 Lut Mergaert (2012), *supra*, note 24.

37 Lut Mergaert (2012), *supra*, note 24.

within DG Justice; a location that does not possess the authority to exercise power over other DGs. Also the robustness, scope and overall coherence of the evaluation system in the Commission contrasts quite sharply with the lack thereof for gender mainstreaming. All in all, the degree of institutionalisation of both approaches is quite different. Considering the significant efforts undertaken to institutionalise evaluation within the Commission,<sup>38</sup> it is perhaps not surprising that the difference is in the advantage of evaluation, despite the constitutionalised status of gender mainstreaming.

## VI. Conclusion: *Ex ante* and *ex post* Evaluations as Two Sides of the Same Coin?

This article brought together two themes of contemporary European governance: evaluation and gender mainstreaming. It has presented an empirical analysis of evaluation texts which accompany EU research policy development through its multiple iterations from FP5 (1998-2002) to Horizon 2020 (2014-2020), with a view to assessing the extent to which *ex ante* and *ex post* evaluations are linked. The analysis returned a number of findings, outlined above: 1) mismatching evaluations; 2) non-linear policy learning; 3) differences between different evaluation types; and 4) the importance of institutional factors. Drawing on these research findings, the conclusion returns to the two research questions that guided analysis. Firstly, how and to what extent are *ex ante* and *ex post* evaluations in EU research policy linked coherently and effectively with respect to gender equality? Secondly, what role has evaluation played in the process of gender mainstreaming in EU research policy?

Beginning with the first, looking across the period of analysis, it cannot be said that *ex ante* and *ex post* evaluations in EU research policy are linked coherently or effectively with respect to gender equality. Whilst the findings do identify some common el-

ements that run throughout most evaluations, and whilst there is some linking between certain evaluations, there is no clear thread of policy learning over time. There is a clear loss in coherence in the gap between *ex post* and *ex ante* evaluations, in addition to evidence of selective coherence. There are a number of factors that undermine a more effective and coherent link between (*ex ante* and *ex post*) evaluations. These are procedural, political and institutional in nature. The procedural factors include, firstly, the mismatch between the *ex post* and *ex ante* evaluations in the policy cycle. This skewed timing is problematic in terms of the use and usability of evaluation outputs.<sup>39</sup> Secondly, addressed as a procedural issue here, there was a particularly striking gap between the *ex post* evaluations and the IIAs in terms of gender equality. This is perhaps because these evaluations developed separately from each other,<sup>40</sup> and is accentuated given the weakness of gender in the IIA. Beyond these procedural challenges and recognised by the Commission itself, political factors have also determined the extent to which gender equality has been addressed as part of EU research policy (text#9, p. 195). This has perhaps been a key factor in explaining the failure to fully mainstream gender-specific evaluation-related exercises in the general process of evaluation, thus severely limiting their formative potential. Finally, institutional challenges, including those associated with a “liquid bureaucracy”,<sup>41</sup> have undermined the continuity of attention to gender mainstreaming over the multiple iterations of the FP.

Turning to the second question regarding the role of evaluation in gender mainstreaming in EU research policy, this role should neither be dismissed nor over-stated. Although complemented with other initiatives (conferences, studies and analyses done by expert groups), the Gender Impact studies completed during FP5 (text#2) played a crucial role in setting the scene for gender mainstreaming in FP6. However, beyond this formative role in the early days of gender mainstreaming in this policy area, the impact of gender-specific evaluations has been limited. Their formative potential has been restricted by limited visibility (see section V, part 3) and selective coherence (see section V, part 2). Analysis also identified some dissonance between how gender equality was addressed in the evaluation exercises and how it played out in the FPs. This was particularly pronounced in the mismatch between gender in the IIA

38 This is neatly described in Steven Højlund (2014), *supra*, note 5.

39 Vítor Oliveira and Paulo Pinho (2011), *supra*, note 2; Raynald Pineault, Paul Lamarche, Matie-Dominique Beaulieu, *et al.*, (2012), *supra*, note 14; Murray Saunders (2012), *supra*, note 14.

40 Vítor Oliveira and Paulo Pinho (2011), *supra*, note 2.

41 Lut Mergaert (2012), *supra*, note 24.

for Horizon 2020 and the provisions for gender mainstreaming in the legal text establishing Horizon 2020 (Regulation (EU) No 1291/2013).

In light of these findings, a number of suggestions are advanced to strengthen the link between *ex ante* and *ex post* evaluations in EU policy-making, with respect to gender equality in particular, as well as more broadly. Firstly, given their primary importance in providing information for *ex ante* evaluations, the formative potential of the FP's interim evaluations should be maximised. Secondly, specifically focused evaluations (such as gender-specific evaluation-related exercises) should be amongst the core of resources used in subsequent general evaluations, and written into the terms of reference of evaluations. This is particularly important due to the vulnerability of gender equality considerations in a complex evaluation process. Thirdly, the status of the EU's constitutionalised commitment to gender equality should be firmly institutionalised throughout the evaluation system; so, gender should be mainstreamed throughout retrospective evaluations as well as the IIA. Finally, horizontal objectives (such as gender equality) should be fixed and assessed over multiple cycles.

## Annex: Selected Texts

**Text#1 (2000):** Five-Year Assessment of the European Union Research and Technological Development Programmes, 1995-1999, Report of the Independent Expert Panel chaired by Joan Majó

**Text#2 (2001):** Commission of the European Communities (2001) "Gender in Research. Gender Impact Assessment of the specific programmes of the Fifth Framework Programme. An overview." Undertaken by external contractors

**Text#3 (2004):** Five-Year Assessment of the European Union Research and Technological Development Programmes, 1999-2003, Full Report, prepared by an external expert group

**Text#4 (2005):** Commission of the European Communities (2005) "Annex to the Proposal for the Council and European Parliament decisions on the 7th Framework Programme (EC and Euratom). Main Report: Overall summary Impact assessment and *ex ante* evaluation", 6 April 2005, SEC(2005) 430

**Text#5 (2008):** Commission of the European Communities (2008) "Gender equality report. Sixth Framework Programme." October 2008, undertaken by the Commission

**Text#6 (2009):** Expert Group on the *ex post* evaluation of the sixth Framework Programmes (2002-2006) (2009) "Evaluation of the Sixth Framework Programmes for Research and Technological Development 2002-2006"

**Text#7 (2009):** European Commission (2009) "Monitoring progress towards Gender Equality in the Sixth Framework Programme, Synthesis Report." May 2009, compiled by the Centre for Strategies and Evaluation Services (CSES) on the basis of contributions by the Gender Monitoring Studies Contractors and the European Commission

**Text#8 (2010):** Interim Evaluation of the Seventh Framework Programme (2007-13), Report of the Expert Group, 12 November 2010, Chair: Rolf Annerberg

**Text#9 (2010):** European Commission (2010) "Stock-taking 10 years of "Women in Science" policy by the European Commission 1999-2009"

**Text#10 (2011):** European Commission (2011) "Horizon 2020 - The Framework Programme for Research and Innovation - Impact Assessment Report", 30 November 2011, SEC(2011) 1427 final